

Section 1.2- Measures of Center and Spread

Measures of Center

Mean

Formula:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

total # of obs. (points to n)
sigma (points to \sum)
obs. (points to X_i)

Symbols:

\bar{X} = sample mean

μ = population mean
mu (points to μ)

Median

Process:

- list data in order (small to large)
- count to the middle observation
- * if you have an even number of observations, average the two middle numbers

$$n = 30$$

$$\bar{X} = 66''$$

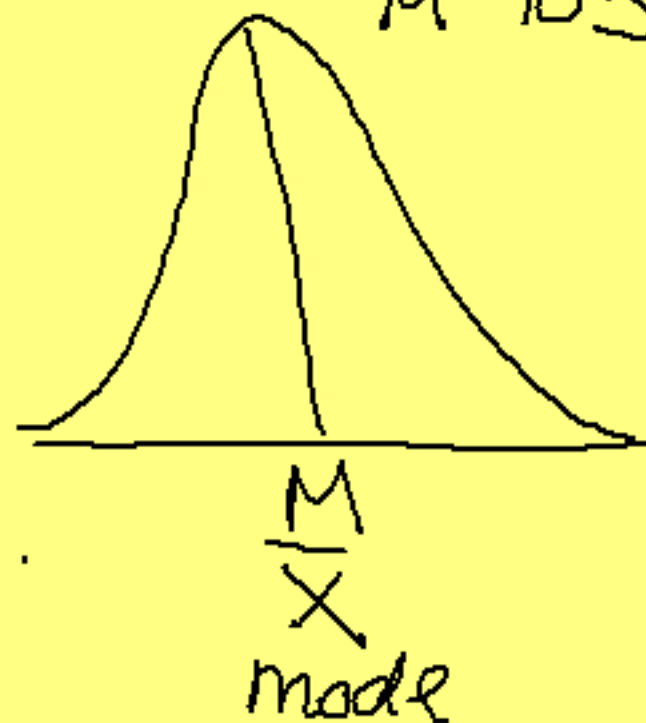
$$\mu = 65''$$

Mean	Median
<ul style="list-style-type: none"> - the average - non-resistant 	<ul style="list-style-type: none"> - the middle # - resistant = not affected by outliers
Skewness	

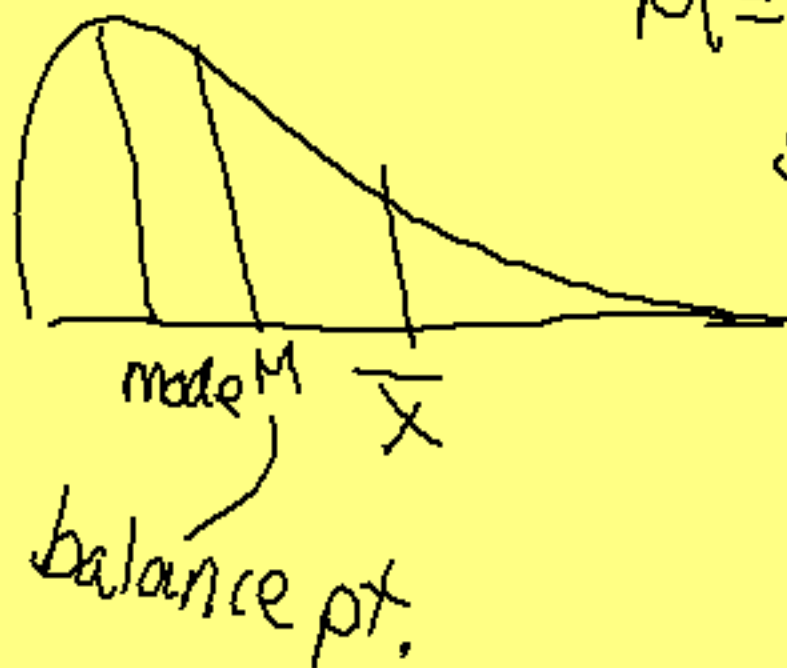
Ex: class heights

Where are the measures of center for different shaped distributions?

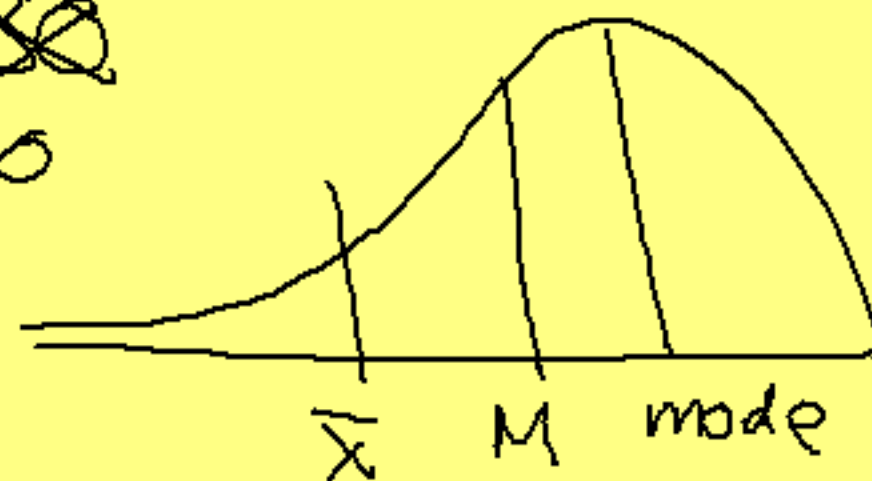
Symmetric $\bar{X} = 100$
 $M = 105$



Right Skewed $\bar{X} = 100$
 $M = 80$
 50



Left Skewed



Measure of Center: MEDIAN

Measure of Spread: 5 # summary

Min, Q1, Median, Q3, Max

**Used for: SKEWED or symmetric distributions
(because Median is resistant)**

Range: full spread of data, min to max, listed as (a,b)

Quartiles: splits the data into 4 equal sections

Q1 M Q3

IQR: Inter-quartile range = $Q3 - Q1 = \#$

- What is the IQR used for?
 - Shows... **middle 50% of the data**
 - Helps... **identify outliers**



IQR Test for possible outliers:

- Find IQR, then multiply by 1.5
- add this to Q3; also subtract this from Q1
- this is the range for "normal" data.
- anything outside this is a possible outlier

+ look @
picture

$$Q_3 = 50 \quad \text{max} = 98 \quad (-22.5, 93.5)$$
$$Q_1 = 21 \quad \text{min} = 10$$

$$IQR = 50 - 21 = 29$$

$$1.5 \times 29 = 43.5$$

$$Q_3 + 43.5 = 93.5$$

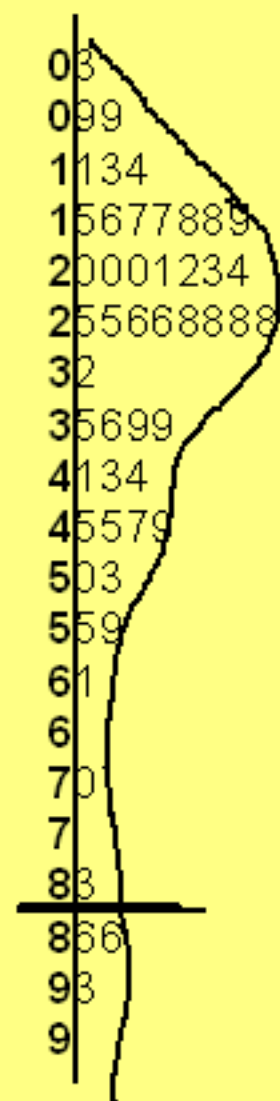
$$Q_1 - 43.5 = -22.5$$



Why is it just for POSSIBLE outliers?

Example: Supermarket Spending- data below is the amount (rounded to the nearest \$) spent in the supermarket. Do a test for possible outliers

Supermarket Spending



Median = 28

Q1 = 19

Q3 = 45

$$IQR = 45 - 19 = 26$$

$$1.5 \times 26 = 39$$

$$Q_3 + 39 = 84$$

$$Q_1 - 39 = -20$$

$$(-20, 84)$$

Percentiles:

- an observation that has a certain percent of the data BELOW it

Ex: 90th percentile

Example: Use the data above to find the percentile for the observation \$24.

total observations = 50

$$\frac{19}{50} = 0.38 \rightarrow 38^{\text{th}} \text{ percentile}$$

Special Percentiles:

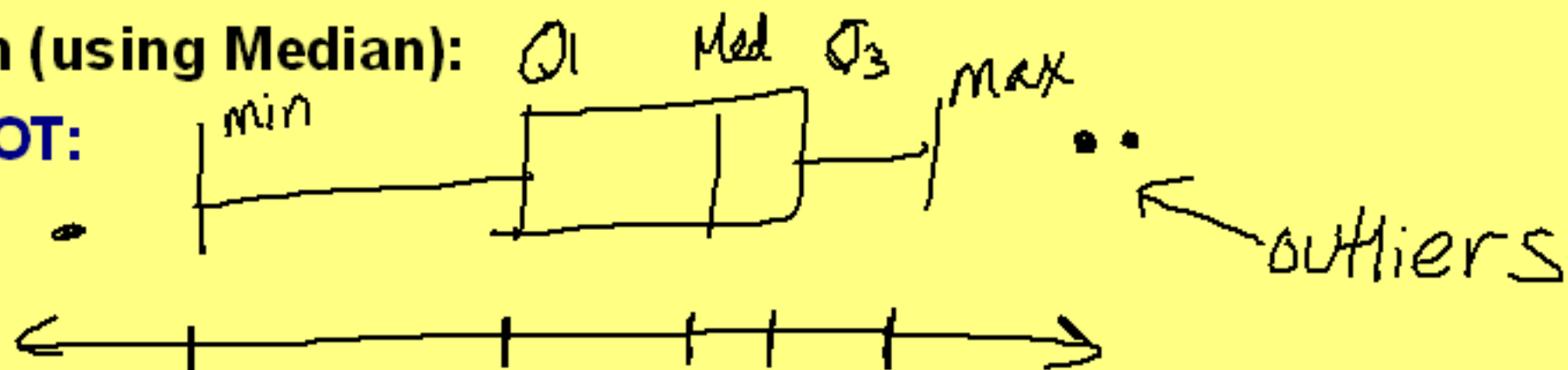
$$Q1 = 25^{\text{th}}$$

$$\text{Median} = 50^{\text{th}}$$

$$Q3 = 75^{\text{th}}$$

Distribution (using Median):

- BOXPLOT:



- side by side boxplots:



On calculator:

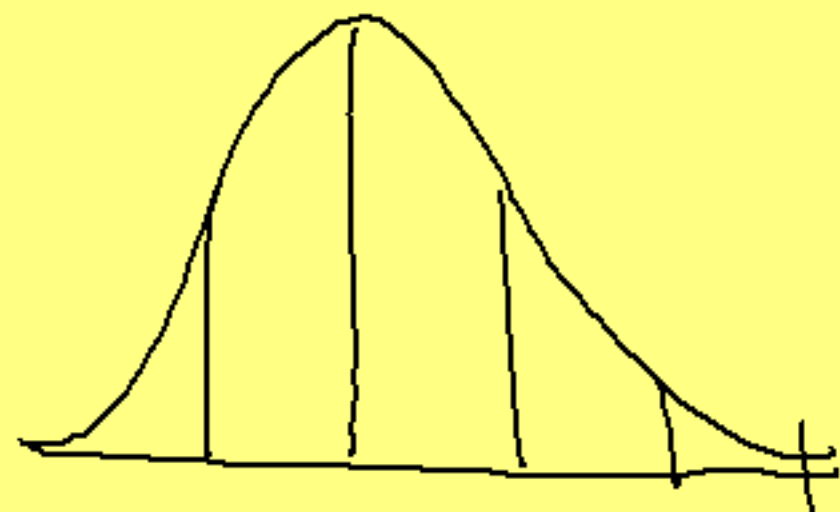
- use the list TEST from yesterday to do a boxplot on the calculator



Measure of Center: MEAN

Measure of Spread: standard deviation

Used for: only symmetric distributions



What IS Standard Deviation?

- **the average distance of the data points from THEIR mean**

- **shows the spread of the data around the mean**

How do we calculate it?

Variance

Symbol:

S^2

σ^2

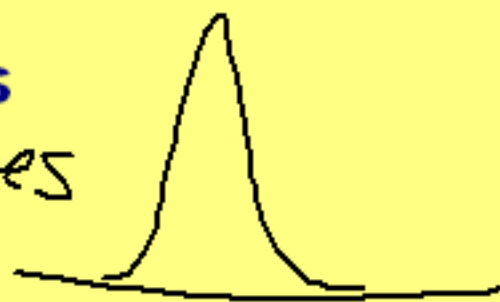
Formula:

$$\frac{\sum (X_i - \bar{X})^2}{n-1}$$

Handwritten notes: "obs" points to X_i , "mean" points to \bar{X} . A bracket under $n-1$ points to the note below.

- Note: $n - 1 =$ **degrees of freedom**

- $\bar{X}_i - \bar{X} =$ **deviations**
differences



Std. Deviation

Symbol:

S

σ

Formula:

Handwritten notes: "Sample" points to S , "pop." points to σ .

$$\sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$$

Handwritten note: "Variance" points to the fraction inside the square root.



Questions: (p. 52-53 in book)

1. Why do we square the deviations?

2. Why do we divide by $n - 1$?

$$n = 10$$
$$\bar{X} = 25$$

3. Why use the std. deviation and not the variance?

in "reg" units

in squared units

Notes on Std. Dev:

- Only use when... **the mean is the chosen measure of center**

- $S = 0$ means... **there is NO spread (points are all the same)**

- Unless there is no spread,... **then $s > 0$**

$$s = 4$$

- As observations become more spread out... **s increases**

$$s = 20$$

- **NON-resistant (because the mean is in the formula)**

Easy way to calculate Std. Deviation:

ON CALCULATOR:

- **STAT -> CALC -> #1:1-VarStats -> ENTER**

- **then put the name of the list**

Please do worksheets:

1.2- Measures of Central Tendency

1.2- Measures of Spread

**Answer keys will be left on the front
table so you can check your work.**